

Article original

Constitution d'une gènothèque destinée à l'identification des génomes mitochondriaux et recherche de nouvelles séquences mitochondriales chez les végétaux supérieurs

B Lejeune¹, A Pereira de Souza¹, N Haouazine¹,
MF Jubier¹, E Besin¹, D Lancelin², G Pelletier²

¹ Université Paris XI, laboratoire de biologie moléculaire végétale associé au CNRS (URA 1128), Bât 430, 91405 Orsay Cedex;

² Institut national de la recherche agronomique, laboratoire de biologie cellulaire, route de St-Cyr, 78026 Versailles Cedex, France

Résumé – Le génome mitochondrial (mt) des plantes supérieures, dont la taille peut varier entre 200 et 560 kb, est composé d'une population hétérogène de molécules de tailles variables pouvant se recombiner grâce à la présence de séquences répétées. Les gènes présents sont constitués de séquences codantes très conservées et peuvent présenter des caractéristiques structurales comme la présence d'introns interrompus conduisant à un épissage en trans. Ces séquences codantes peuvent être utilisées comme sondes pour caractériser le polymorphisme entre variétés ou espèces. La recherche de nouvelles séquences codantes utilisables nous a amenés à identifier et caractériser 2 gènes mt nouveaux, nad5 et nad6.

mitochondrie / RFLP / nad5 / nad6 / épissage en trans

Summary – The use of a mitochondrial (mt) gene library for characterization of mt genomes and the search for new mt genes in higher plants. The mt genome of higher plants (size varying from 200 to 560 kb according to the species) consists of a heterogeneous population of molecules resulting from intra- or inter-molecular recombination. Besides the highly conserved coding sequences of mt genes, other sequences, for example, promiscuous chloroplast sequences, may be found. The genes may be interrupted by introns; in some cases, one or more introns have been split, leading to the scattering of the gene segments to remote parts of the genome; this structure requires trans-splicing steps for the mature messenger RNA to be obtained. Over 25 genes have been identified so far that are usable, as universal polymorphism markers in RFLP (restriction fragment length polymorphism) characterization of mt genomes due to sequence conservation. During the search for new genes we identified 2 new coding sequences, nad5 and nad6; nad6 consists of a single

reading frame whereas nad5 is a split gene whose 3 independent segments are independently transcribed.

mitochondria / RFLP / nad5 / nad6 / trans-splicing

LE GÉNOME MITOCHONDRIAL DES PLANTES SUPÉRIEURES

Organisation

Le génome mitochondrial (mt) des plantes supérieures peut être représenté par une structure modèle consistant en un cercle-maître présentant des régions répétées plus ou moins longues dont la recombinaison engendre une population hétérogène de molécules subgénomiques (Quétier *et al*, 1985). Il s'agit donc d'un génome structurellement hétérogène, dans lequel la même séquence mt peut, pour la même plante, être trouvée simultanément dans des environnements moléculaires différents. Les cercles-maître ont une taille physique (exprimant donc la complexité du génome) comprise entre 200 et 600 kb suivant les espèces, ce qui fait du génome mt des plantes supérieures un des plus grands génomes mt connus.

Les nombreux remaniements structuraux possibles (Small *et al*, 1989) conduisent, même entre espèces très proches, à des génomes mt de contenu analogue mais d'organisation complètement différente dont l'histoire mouvementée peut être reconstituée dans certains cas (Palmer, 1988).

Contenu informationnel

Le génome mt des plantes code pour un certain nombre de constituants de l'appareil de traduction mt (ARN ribosomiaux et protéines ribosomiales, ARN de transfert) et de complexes de la membrane interne assurant les transferts d'électrons (complexes I, III et IV) et la production d'ATP (ATP synthase). Plus de 25 gènes ont ainsi été identifiés pour l'instant.

On trouve aussi d'autres séquences dont l'origine peut parfois être retrouvée (chloroplastique ou nucléaire) et dont la nature varie suivant l'espèce. Il ne semble pas que ces séquences (sauf quelques tARN chloroplastiques) soient utilisées par la plante.

En résumé, le génome mt des plantes peut être considéré comme une mosaïque entremêlant un fonds commun de gènes avec des séquences diverses et variables suivant les espèces, les arrangements de l'ensemble étant propres à une espèce donnée et stables dans cette espèce.

Quatre phénomènes importants dans la structure, l'expression et la localisation de certains de ces gènes ont été mis en évidence depuis 1989; ce sont, chronologiquement :

– la *très grande conservation de la séquence d'un même gène* chez des plantes supérieures éloignées évolutivement : les comparaisons de séquences montrent que les taux de similarité sont pratiquement toujours supérieurs à 90% ;

- l'*édition des ARN* : il est maintenant établi que les transcrits primaires des gènes mt végétaux codant pour des protéines subissent une «édition» ultérieure consistant en des changements spécifiques de certains C en U qui, en modifiant les codons, entraînent des changements d'acides aminés dans la séquence protéique (Walbot, 1991);
- l'*épissage en trans* : les gènes mt des plantes supérieures peuvent comprendre des introns de classe II. Certains de ces gènes composites ont eu, à la suite de remaniements, leurs introns scindés; les morceaux de gènes résultant de cette scission ont été dispersés dans le génome mt où ils sont transcrits séparément; la réalisation d'un ARN messager mature dépend donc de transcriptions indépendantes suivies de la reconnaissance des différents transcrits par «épissage en trans» (Bonen, 1993); ce phénomène, décrit originellement pour des gènes chloroplastiques, a été retrouvé dans les mitochondries pour les gènes *nad1*, *nad2* et *nad5* (voir ci-dessous);
- le *transfert de gènes mitochondriaux vers le noyau* : il existe actuellement quelques exemples de gènes mt dont des copies sont présentes dans le génome nucléaire chez la même plante. Ces exemples sont la preuve d'un passage progressif de gènes mt vers le noyau et témoignent du devenir des génomes endosymbiotiques originaux (Nugent et Palmer, 1991).

LES STRATÉGIES DE MISE EN ÉVIDENCE DE GÈNES

L'identification de ces gènes repose sur 3 stratégies :

- la *détection par hybridation avec des sondes hétérologues* provenant d'autres génomes mt : ce procédé a permis la détection de certains gènes à partir de sondes de champignons ou d'animaux; il souffre de 2 limitations : d'une part les *divergences entre séquences nucléotidiques* pour un même gène qui, si elles sont trop marquées, ne permettent plus d'obtenir de signaux significatifs dans les expériences d'hybridation; d'autre part l'existence, dans les génomes mt des plantes, de *gènes n'ayant pas d'équivalent dans les mitochondries d'autres groupes* d'organismes;
- l'*utilisation des transcrits mitochondriaux* pour l'identification des gènes dont ils dérivent : cette stratégie, qui ne fait pas appel à un génome étranger, procède par *repérage des régions au moyen d'hybridations de type Northern* ou par *la construction de banques de ADNc*; les cadres de lecture détectés sont identifiés par comparaison avec les banques de données;
- le *séquençage systématique de l'ADN mt et la recherche de cadres de lecture* : ce système, appliqué à un génome entier, permet de mettre en évidence de nombreux cadres de lecture potentiels; cela a été réalisé pour le génome mt de *Marchantia polymorpha* (qui est une Bryophyte) et le sera prochainement pour le riz et *Arabidopsis*.

LA MISE EN ÉVIDENCE DES SÉQUENCES CONSERVÉES ET L'IDENTIFICATION DE NOUVEAUX GÈNES

La stratégie

Le grand degré de conservation des séquences des gènes mt des plantes supérieures a été rapporté ci-dessus; il se manifeste même entre plantes appartenant à des familles évolutivement distantes. Nous avons mentionné également le fonds commun que constituait l'ensemble des gènes mt des angiospermes à l'intérieur de génomes de compositions par ailleurs hétérogènes. Ces 2 faits nous ont conduits à mettre au point une stratégie de détection de gènes mt basée sur la mise en évidence de ce fonds commun de séquences conservées. Différentes stratégies sont envisageables. Nous avons choisi d'en utiliser 2.

Dans la première, un génome mt entier (celui de *Brassica oleracea*) a été marqué radioactivement et a servi de sonde envers un génome mt cible, celui du blé, utilisé sous forme d'une banque de cosmides connus et cartographiés.

Dans une seconde série d'expériences, des fragments d'environ 3 kb issus des cosmides de la banque mt de blé ont été utilisés comme sondes dans des hybridations sur taches (Dot Blot) d'ADN mt de 6 dicotylédones différentes. Les fragments sondes donnant un signal positif sur ces 6 ADN ont été sélectionnés. Cette expérience permet d'estimer à 120-150 kb la longueur du fonds commun de séquences conservées entre génomes mt.

Les résultats

Parmi les séquences révélées dans le génome mt de blé par ces hybridations croisées, nous avons sélectionné celles qui ne correspondaient pas à des gènes déjà identifiés et qui étaient transcrites. Deux d'entre elles ont fait l'objet d'un travail d'identification approfondi et ont conduit à la caractérisation des gènes *nad5* et *nad6*.

Le gène nad5

Le gène *nad5* appartient à la catégorie des gènes éclatés comme *nad1* ou *nad2*, dont l'expression requiert une ou plusieurs étapes d'épissage en *trans*. Dans le cas de *nad5*, le gène est constitué de 5 exons groupés en 3 régions séparées du génome.

À cette organisation multipartite du gène *nad5*, retrouvée chez toutes les Angiospermes testées, correspond une transcription très complexe résultant des épissages en *trans* et en *cis* des 3 transcrits (Pereira de Souza *et al*, 1991).

Le gène nad6

La séquence du deuxième fragment choisi a révélé un cadre de lecture ouvert. Malgré des difficultés dues aux divergences de séquences pour un gène semblant évoluer rapidement, nous avons identifié le gène *nad6*, présent chez toutes les Angiospermes testées. Sa localisation est variable suivant les espèces. Dans le cas du blé, des réarrangements l'ont intégré à l'un des introns scindés du gène *nad1* (ce qui conditionne sa transcription à celle de ce dernier) alors que chez le maïs il est situé près du gène *atp6* (Haouazine *et al*, 1992).

CONCLUSION

Le polymorphisme du génome mitochondrial des plantes supérieures et sa mise en évidence par RFLP (polymorphisme de longueur de fragment de restriction)

Nous avons mentionné dès le début de cette revue le caractère hétérogène de la structure du génome mt des plantes supérieures qui se manifeste par des profils de restriction complexes, comportant de nombreuses bandes de stoechiométrie variable et d'interprétation parfois difficile. La comparaison de profils de restriction peut néanmoins constituer un moyen d'analyse du polymorphisme.

Une deuxième approche utilise les gènes mt communs aux plantes supérieures. La grande conservation des gènes fait de ces séquences d'excellentes sondes universelles pour mettre en évidence le polymorphisme du type RFLP dans l'arrangement ou l'intégration d'une séquence donnée. Cette approche a déjà été très utilisée et, comme toute technique basée sur le RFLP, est d'autant plus efficace que l'on dispose de davantage de sondes. L'identification de nouveaux gènes mt a fourni et fournira en parallèle les nouveaux marqueurs RFLP nécessaires à la caractérisation fine des génomes mt des plantes supérieures.

REMERCIEMENTS

Ce travail a bénéficié de l'aide du ministère de la Recherche et de la Technologie (action 89 G 0453).

RÉFÉRENCES

- Bonen L (1993) Trans-splicing of pre-mRNA in plants, animals, and protists. *FASEB J* 7, 40-46
- Haouazine N, Pereira de Souza A, Jubier MF, Lancelin D, Delcher E, Lejeune B (1992) The wheat mitochondrial genome contains an ORF showing sequence homology to the gene encoding the subunit 6 of the NADH-ubiquinone oxidoreductase. *Plant Mol Biol* 20, 395-404
- Nugent JM, Palmer JD (1991) RNA mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. *Cell* 66, 473-481
- Palmer JD (1988) Contrasting modes and tempos of genome evolution in land plant organelles. *Trends Genet* 6, 115-120
- Pereira de Souza A, Jubier MF, Delcher E, Lancelin D, Lejeune B (1991) A trans-splicing model for the expression of the tripartite *nad5* gene in wheat and maize mitochondria. *Plant Cell* 3, 1363-1378
- Quétier F, Lejeune B, Delorme G, Falconet D, Jubier MF (1985) Molecular organization and expression of the mitochondrial genome. In: *Higher Plant Cell Respiration* (Douce R, Day DA, eds). Springer Verlag, Berlin, 18, 25-36
- Small I, Suffolk R, Leaver CJ (1989) Evolution of plant mitochondrial genomes via substoichiometric intermediates. *Cell* 58, 69-76
- Walbot V (1991) RNA editing fixes problems in plant mitochondrial transcripts. *Trends Genet* 7, 37-39